Camera Brand Congruence in the Flickr Social Graph

Adish Singla
Ecole Polytechnique Fédérale de Lausanne
adish.singla@epfl.ch

Ingmar Weber
Ecole Polytechnique Fédérale de Lausanne
ingmar.weber@epfl.ch

ABSTRACT

Given that my friends on Flickr use cameras of brand X, am I more likely to also use a camera of brand X? Given that one of these friends changes her brand, am I likely to do the same? These are the kind of questions addressed in this work. Direct applications involve personalized advertising in social networks.

For our study we crawled a complete connected component of the Flickr friendship graph with a total of 67M edges and 3.9M users. Camera brands and models were assigned to users and time slots according to the model specific meta data pertaining to their images taken during these time slots. Similarly, we used, where provided in a user's profile, information about a user's geographic location and the groups joined on Flickr.

Our main findings are the following. First, a pair of friends on Flickr has a significantly higher probability of being congruent, i.e., using the same brand, compared to two random users (27% vs. 19%). Second, the degree of congruence goes up for pairs of friends (i) in the same country (29%), (ii) who both only have very few friends (30%), and (iii) with a very high cliqueness[1] (38%). Third, given that a user changes her camera model between March-May 2007 and March-May 2008, high cliqueness friends are more likely than random users to do the same (54% vs. 48%). Fourth, users using high-end cameras[2] are far more loyal to their brand than users using point-and-shoot cameras, with a probability of staying with the same brand of 60% vs 33%, given that a new camera is bought. Fifth, these "expert" users' brand congruence reaches 66% (!) for high cliqueness friends.

To the best of our knowledge this is the first time that the phenomenon of brand congruence is studied for hundreds of thousands of users and over a period of two years.

[1]The formal definition is given in Section 4.3.
[2]Here, we looked at digital single-lens reflex cameras.

WSDM '09 Barcelona, Spain

Categories and Subject Descriptors

H.1.1 [**Models and Principles**]: Systems and Information Theory

General Terms

Human Factors, Experimentation

Keywords

Flickr, brand congruence, brand loyalty, social network

1. INTRODUCTION

How much information about personal brand preferences can be derived from knowledge about the brand preferences of one's friends? Which aspects affect the probability to have similar camera preferences to my friends? Are there factors which are crucial when it comes to brand loyalty? These are the kinds of questions we study in this work in the setting of the Flickr[3] online photo sharing site.

Understanding how friends, either in the settings of social networks or defined through personal contacts, affect our decision regarding the purchase of any kind of item, is interesting for a number of reasons. First, it is interesting to see if there is any correlation at all between an online relation (a friendship link on Flickr) and an offline property (the camera brand owned by a user). Second, there are clearly potential applications for more effective, targeted and personalized marketing campaigns, especially in social networks. Third, it makes it possible to quantify the affect of sociological phenomena such as "peer pressure" concerning purchase decisions.

The availability of the Flickr data offers the possibility to address such issues, not just for a few hundred people via personal surveys, but for millions of users and using information for several years. The main observation required to perform such an investigation lies in the fact that the majority of pictures uploaded to Flickr come with machine related meta data, which contains information about the brand of the camera used, the exact model specification and also the date when the image was taken. This then allows us to assign a brand and a model to a user for a given period of time. The exact details will be explained in Section 3. Furthermore, Flickr profiles (optionally) contain the geographic location of a user and indications of the degree of online activity, such as the number of images uploaded by a user. All this information creates a plethora of dimensions to explore.

[3]`http://www.flickr.com`

The main focus of our study are effects which influence the probability that two users own cameras of the same brand. We chose to focus on brands rather than individual models, as models are more volatile and even if one friend convinces another after three months of the satisfaction with her *particular* camera, there will probably already be a new, similar model by the same brand on the market. Note that Flickr displays information about the camera used to take a picture on the picture's main page under "Additional information", so that even users without any direct personal contact can take note of this.

Our main findings, explained in detail in Section 4, are the following. First, a pair of friends on Flickr has a significantly higher probability of being congruent, i.e., using the same brand, than two random users (27% vs. 19%), where these numbers refer to the time period of March to May 2008. Second, this effect can *not* be solely explained by geographical factors. Friends are more likely to be in the same country, but even random users in the same country are still less likely (23%) to be congruent than friends, and in particular than friends in the same country (29%). Third, the degree of congruence goes further up for pairs of friends (i) who both only have very few friends (30%), and (ii) with a very high degree of cliqueness[4] (38%). Fourth, given that a user changes her camera model between March-May 2007 and March-May 2008, high cliqueness friends are more likely than random users to do the same (54% vs. 48%), and 38% of the high cliqueness friends who do change their model in the same period, change to a camera of the same brand. Fifth, users using high-end cameras cameras are far more loyal to their brand than users using point-and-shoot cameras, with a probability of staying with the same brand of 60% vs. 33%, given that a new camera is bought. In both cases, users are more likely to buy the same brand again than it is that a random user would buy the particular brand. Sixth, for users of high-end cameras the cliqueness is most relevant, increasing the baseline probability for two friends using such cameras to be congruent from 47% to 66%.

All the techniques used and the factors investigated are not limited to Flickr in any way. Similar studies could be easily done for other product-related social networks. In particular, there are several sites related to fashion[5], to mobile devices[6], to cars[7], to books[8] and probably to other products as well. In some of these cases, the site might not directly offer friendship links, but such connections could be implicitly deduced from comments left by users.

The rest of this paper is organized as follows. In Section 2 we discuss work related to our analysis. Section 3 gives details about our data set and how it was obtained. Section 4 contains the actual results of our work. This section is split into three parts. First, in Subsection 4.1, we will present the techniques used by us. Then, in Subsection 4.2, we focus on (i) static analysis for a single time period and on (ii) more straightforward properties to investigate. Subsection 4.3 then goes beyond this by looking at the evolution over time and at more subtle features. Finally, in Section 5 we discuss possible extensions of our work and give a summary of our main findings.

[4]This term refers to the overlap between two sets of friends and is formally defined in Section 4.3.

[5]`http://www.fashmatch.com/`, `http://shareyourlook.com/`

[6]`http://pdaphonehome.com/`,`http://www.modaco.com/`

[7]`http://www.carspace.com/`, `http://www.carsablanca.de/`

[8]`http://www.shelfari.com/`, `http://www.librarything.com/`

2. RELATED WORK

As Flickr has a very rich and interesting set of data and as this data is accessible via a public API, it has been used for several studies. Among other things investigated, people have looked at general graph properties [11], and they have extensively used the *tags*. The usage pattern of tags was investigated in [10], place and event event names were automatically derived in [12], and the Flickr tags were used to design and evaluate tag recommendation systems [15] [5].

The study of user behavior in large, online social networks is by no means new. Most relevant to our study is the work in [16], as they also consider the degree of similarity between two users who are connected. In their setting, the network studied is the MSN network and the links they used correspond to chat sessions between users. Similarity between two users is measured with respect to web queries made and with respect to personal information such as age or location. They do not consider any product-related information, nor do they consider any graph related properties, such as cliqueness. The same data set was also used in [8], where the focus is on the actual instant messaging behavior. That is, they investigate which factors influence the number and length of conversations. They also consider general graph properties and verify the "six degrees of separation". In [11] the focus is on properties related to link distributions in several large-scale online social networks. In particular, they investigate inlink and outlink distributions for Flickr. The work in [1] is similar to this, but uses different data sets also includes the factor of evolution over time. More related to our study is the work presented in [3]. Here, the authors look at factors that govern the growth of communities. Their analysis uses LiveJournal (with its communities) and DBLP (with scientific conferences). In their setting, the single most influential feature turned out to be the "proportion of friends in a community who are friends with each other", which is similar to our notion of "cliqueness". Although we are unsure if their findings will apply to setting where (i) a user can only join a *single* community and (ii) there is a financial cost associated with joining a community, we plan to use a framework related to theirs in the future to investigate the evolution over time more closely.

There is also work related to brand congruence in social networks and to "viral marketing". First, Flickr itself offers rudimentary, aggregate statistics about the cameras used by its users[9]. Then, there has been previous work on mining social networks for targeted advertising. In [17], the authors look at a social network derived from email messages between 427 faculty members of a university. The products they study are books, where the data refers to book loans from the library. The "brands" of interest are topical categories. The authors then find that highly cliqued[10] groups of users, supposedly loosely corresponding to faculties, are more likely to borrow books on the same category. Though similar in spirit, our study differs in size, an actual use of

[9]`http://www.flickr.com/cameras/brands/`

[10]Our definition of "cliqueness" is identical to their definition of "cohesion".

product-related information, the techniques used and the thorough analysis of various factors influencing the strength of links.

Related to but different from our work are studies looking at recommendation networks such Epinions[11]. Here users explicitly rate certain products, they often have an explicit network of trusted users (concerning product reviews) and they often have the explicit possibility to recommend a product to other users. [14] explicitly raises the question how to identify the "best" users to target for viral marketing. To answer this question, given certain game theoretical modeling assumptions, both a hardness result and an approximation algorithms are presented in [6]. In [9], authors discuss the algorithms to enumerate the recommendation patterns. [7] presents simple stochastic models to explain the patterns and size of recommendation cascades. These works are also more concerned with modeling the evolution over time, assuming that the strength of influence between pairs of nodes is *known*. So, in a sense, our work is somewhat orthogonal to this group of work, as we are mostly interested in understanding which factors govern the strength of links.

An interesting, very detailed study on brand congruence in real-life social networks was investigated in [13]. Here, the "nodes" of the social network were 49 members of a sorority at a university. Link types ranged from "roommate", via "sharing a bathroom" to "joint sports activities". Product types ranged from shampoo, via tv shows to pizza. Although the data set was very small and sparse, the authors could find significant effects for, e.g., room neighbors sharing pizza preferences or people sharing a bathroom using the same brand of shampoo. Though this kind of detailed study via surveys certainly does help to identify certain effects, it has clear scalability problems when it comes to the analysis of global networks with hundreds of thousands of users.

There has also been work on the study of explicit brand communities, that is, clubs centered around a certain product brand and often directly sponsored by the corresponding corporation. One such study, focusing on car clubs, is presented in [2]. The authors show that, e.g., the most active group members are those who feel most positively towards the brand. However, it is also argued that simply trying to market the club membership, in an effort to then boost sales for the new members, will generally not work as a positive attitude has to precede an active role in the club. Finally, there is also the aspect of *brand loyalty*, both in the sense of purchase and attitudinal loyalty. The relevance of psychological factors such as brand trust and brand affect was investigated in [4]. We also look at brand loyalty, but our focus is more on user types. In particular, we looked at whether "expert" users are more or less loyal to their brands than other users.

3. DATA SET

All of our data was obtained using the public Flickr API[12]. Basic statistics about the data are given in Table 1. We crawled a complete connected component of the Flickr friendship graph, starting with a very active user with over 100 contacts[13]. As part of this crawl, we downloaded the complete list of public photos for each user. In this initial phase, we obtained a graph containing 3.9M users and 67M friendship edges, as well as a list of 500M images with their "date taken" information. Note that, while it is possible to get the list of (public) pictures of a user, along with the corresponding "date taken" information, via a single API call, other information has to be obtained on a per-image basis.

We now pruned this initial set of users, as we were only interested in users for which we could extract brand information, which in turn was derived from information for uploaded images. Therefore, we removed all users from the data set who had not uploaded any public pictures. This left us with 2.1M users and 44M friendship links. Note that friendship edges in Flickr are not necessarily symmetric, as adding a friend does not require authorization. This means that user A can have B as a friend, without user B having A as a friend. Of the 44M links, 30M belong to pairs of bidirectional links and the remaining 14M links are not reciprocated by the other person.

After we had obtained the complete list of users, we then went on to obtain additional information for each user of interest. In particular, where provided, we obtained the country of the user, as provided in her user profile. Here, quite a bit of care had to be taken as the location could be entered in a free text format so that all of "California", "San Fransisco", "USA", "America" and even "Canada's neighbor" needed to be mapped to the same, unique country ID. Out of the 2.1M users investigated, 581K specified some country and 519K of these cases were mapped to a valid country ID[14]. Furthermore, for each user we obtained the groups joined on Flickr. Groups on Flickr represent communities of users sharing common interest, usually related to photography. All of this data is (i) static in the sense that it does not relate to a particular period of time and (ii) not directly connected to the brand or model information, which is our focus of interest.

The key observation that allowed us to obtain brand and model information, is that when images are uploaded to Flickr, the so-called Exif[15] meta data is usually preserved. This meta data contains information about the manufacturer, the model, as well as other information related to resolution or exposure time, which was not used by us. It is also available on a picture's main page on Flickr under the header "Additional information". Given this Exif data, we then assigned brands and models to a user for a particular period of time as follows. First, we chose three time slots to investigate, namely, the period of March 1 to May 31 for the years 2006, 2007, and 2008. Then, for each of these three time slots, we tried to obtain the Exif data for up to 10 public images for each of the 2.1M users of interest. The reason that we used only 10 images per user is that for each Exif data a separate call to the API had to be issued and one such call took, roughly, 1 second. Obtaining this data for *all* of the public 500M images discovered during the crawl, would not have been feasible in an acceptable time frame.

If a user had uploaded more than 10 images for a time period of interest, we first tried to obtain the relevant Exif information for the image closest to the center of the slot (April 15) and then worked our way towards the ends of the interval in a symmetric fashion. In the vast majority of the

[11] `http://www.epinions.com`

[12] `http://www.flickr.com/services/api/`

[13] `http://www.flickr.com/people/acastellano/`

[14] Unmapped cases include, e.g., "land of Putin" or "I am Italian".

[15] `http://en.wikipedia.org/wiki/Exchangeable_image_file_format`

cases (94%), all the meta data obtained (if there was any) for a user in a time slot was consistent in the sense that only a single brand occurred. For the cases where this was not the case, we used a simple majority voting scheme and assigned the strongest brand within a time slot to each user, where ties were broken at random. In the same manner, users were assigned a model for each time slot, if at least one of the (up to) 10 images contained a valid model information. Mapping brands to unique IDs again required some manual labor. For example, we took care to map "Minolta", "Konika" and "Konica" to the same unique brand ID. As for mapping camera models to IDs, we used the list of 1,785 cameras available at `http://www.flickr.com/cameras/`. This list also contains information about the camera's type, in particular whether it is a point-and-shoot (P&S) or a digital single-lense reflex (DSLR) camera. We again tried our best to ensure that different camera names (such as "Maxxum 7D" and "Dynax 7D") referring to the same model were mapped to the same ID. Cameras not included in the online list, apparently rather old ones, were not mapped to model IDs.

Table 1: Some basic numbers describing our data set. More brands and countries were present, but they were too insignificant to be picked up by our hand-crafted mappings to IDs

Number of users	
- before pruning	3.9M
- after pruning	2.1M
Number of edges	
- before pruning	67M
- after pruning	44M

Number of ...	
brands	96
models	1785
countries	168
groups	203K
users w. country	510K
users w. group	850K

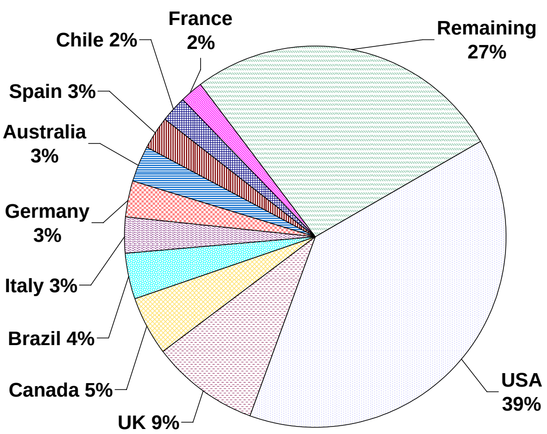

Figure 1: User distribution across countries for the 24% (out of 2.1M) of users who specified valid country information.

It should be noted that although we have time-related information about the users' camera brands, the underlying friendship network, the geographical information and the groups information is only for June 2008, the time of the crawl. Hence, the fact that two users are friends in our data set does not necessarily mean they were friends in 2006. However, we do assume that the friendship network in 2008 is still a good approximation of the network in 2006, and so we also include results for this period. But our focus is always on the most current time slot. Also note that although 2.1M users in our crawl have at least one public image, only 1.2M of them uploaded a picture with valid brand information during our time periods of interest.

Table 2: An "active" user is one for whom we could successfully obtain brand information. Numbers in parentheses refer to the number of users with valid model information. The reason that the numbers for the slot in 2008 is *lower* than for 2007 is probably that we obtained the data in Mid-June, when not all users had uploaded their pictures for the corresponding period yet. 152K users were active in all three time periods (68K with country information) and 1.2M were active during at least one time period (326K with country information)

	Mar-May06	*Mar-May07*	*Mar-May08*
Active users	470K (350K)	670K (520K)	630K (500K)
... w. country	160K (118K)	210K (164K)	200K (159K)

4. RESULTS

Here, we present the results of our analysis. First, in Section 4.1 we introduce the basic technique of analysis which was used for most of our study. Sections 4.2 and 4.3 then present our main findings, ranging from rather basic brand congruence analysis to a more advanced analysis of how camera changes of one user affect her friends.

4.1 Techniques

Our basic technique of data analysis, also used in [16], is the following: We consider pairs of users of a certain type and measure the percentage of them sharing the same camera brand, an event we refer to as *brand congruence*, or simply as being congruent. E.g., we look at pairs of random users and compare their probability of brand congruence (as a baseline) to the probability for pairs of friends. We then consider more and more relevant conditions for possible pairs of users, such as whether they are in the same country, whether they share many common friends or whether they both use high-end cameras.

We evaluated all of these numbers for the full list of 44M friendship links. However, due to the obvious problems of scalability, we did not compute these numbers for *all* 4.4 trillion (!) possible pairs of random users. Here, we once sampled uniformly at random a set of 44M random pairs, irrespective of any existing friendship links, and then computed all the relevant properties for this collection of pairs. No "self-links" were allowed for this. The 44M pairs were then further conditioned to, e.g., only make statements about pairs of random users in the same country. Both for pairs of friends and for the pairs of random users, we always include the absolute numbers of pairs which have a certain property, along with the percentage of these which are congruent. Note that for certain statistics this sampling was not required. E.g., if 60% of users use brand A and 40% brand B, then one can directly compute that 36%+16%=52% of all user pairs will be congruent. Whereever such a closed-form solution was possible, the results agreed for at least three significant digits.

In Section 4.3 we make use of a more fine-grained analysis, where we track individual users across two time periods and then make statements about their probability (or the probability of their friends) to change their brand or their camera model, given certain properties.

4.2 Basic Brand Analysis

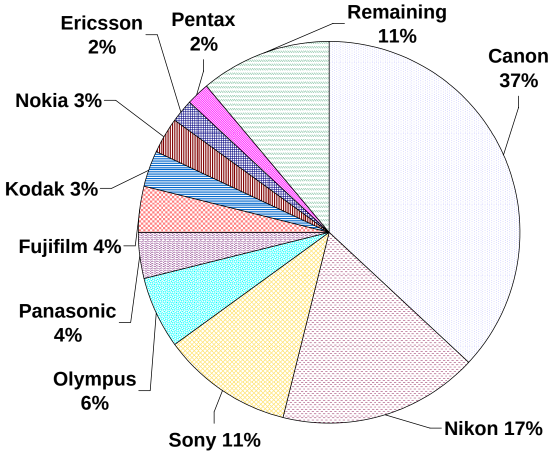

Figure 2: Brand distribution in 2008 for the 200K users who uploaded a picture taken between March and May with a valid brand in the Exif data.

Differences between pairs of friends and random users. As a baseline experiment, to compare effects of various factors against, we measured the probability of brand congruence between any two random users in a given timeslot. Table 3 gives the results for the three time slots. To see a first relevance of the friendship factor when it comes to brand congruence, we measured the probability of congruence on our Flickr friendship network in a given timeslot. Table 4 gives the results for the three time slots.

Comparing Table 3 and Table 4, we see an increase of 80% in the probability of sharing the same brand in the friendship graph as compared to random pairs. This result is observed for all the timeslots. However, it could well be that two friends are simply more likely to be in the same geographical area and that the higher degree of congruence can be solely explained by geographical changes of the predominant brand. Our next experiment will show that this is not the underlying reason for the observed effect.

Table 3: Probability that a pair of two *random* users will share the same camera brand. Here and throughout the paper the number in parentheses give the absolute numbers of pairs used

Mar-May06	*Mar-May07*	*Mar-May08*
0.16 (2.0M)	0.17 (4.2M)	0.19 (3.7M)

Table 4: Probability that two *friends* share the same camera brand. These numbers should be compared with the results in Table 3

Mar-May06	*Mar-May07*	*Mar-May08*
0.22 (5.9M)	0.25 (11M)	0.27 (11M)

Random pairs with country conditions. To see if the underlying explanatory factor is the common country of two friends, we carried out experiments, where we split pairs of random users into two classes, depending on whether they come from the same country or not. Only users who had provided valid information concerning their country were used for this experiment.

Table 5 gives the probabilities that we get by conditioning on random user pairs in the two ways mentioned above. Note that this is *not* a proper breakdown of the results in Table 3, as the majority of users do not have any country information. These users are still included in the previous results, but are not used for the country-related analysis. The fact that conditioning on the same country leads to significantly higher congruence probabilities (see Table 3 for comparison) means that the higher congruence for friendship links can at least partly be explained by regional congruence. However, (i) the congruence probabilities are still *lower* than for two random friends and (ii) the congruence among friends increases further, when we also condition on both users having the same country (and a valid country to begin with), as shown in Table 6.

Table 5: Probability of congruence for *random* pairs, depending on whether they are in the same country (sc) or in different countries (dc)

	Mar-May06	*Mar-May07*	*Mar-May08*
same country	0.19 (64K)	0.22 (70K)	0.23 (44K)
diff. country	0.16 (187K)	0.18 (336K)	0.19 (301K)

Friendship graph with country conditions. As for two random users, we also split the results for two friends, according to whether they come from the same or different countries. Note that even two friends from *different countries* are still more likely to be congruent than two random users from the same country. This is a first indication of the relevance of friendship which will be investigated further in the next section. Figure 3 gives a summary of the results so far.

Table 6: Probability of congruence for *friends*, depending on whether they are in the same country (sc) or in different countries (dc)

	Mar-May06	*Mar-May07*	*Mar-May08*
same country	0.24 (908K)	0.27 (1.4M)	0.29 (1.5M)
diff. country	0.21 (1.3M)	0.24 (2.1M)	0.28 (2.1M)

4.3 Advanced Brand Analysis

In the previous section we saw that friendship information is clearly relevant for brand congruence. In this section we dig deeper and try to find out "which kind of friendships" matter, "which other factors play a crucial role" and "how things evolve with time".

Varying degree of groups similarity. One attempt to measure the closeness of two friends, is to find out if they share common interests. Though lists of interests are not available on Flickr, the groups joined provide some indication to the interests held by a particular user. Therefore, one could hypothesize that two friends who joined many of the same groups are "closer" in some sense and should be more congruent. To test this hypothesis, we classified the links

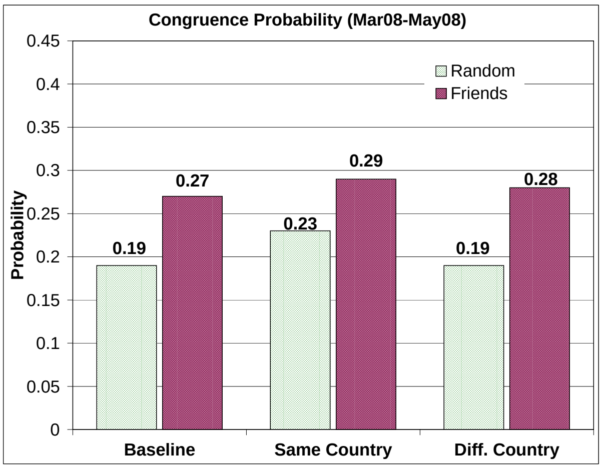

Figure 3: Histogram summarizing the results of our basic analysis for the time period March-May 2008. Note that two friends who provided valid country information, even if they come from different countries, are still more likely to be congruent than two arbitrary friends.

between two friends according to the Jacquard coefficient of the groups joined. This coefficient is defined as $G_J(X,Y) = |G(X) \cap G(Y)|/|G(X) \cup G(Y)|$, where $G(X)$ denotes the set joined by a user X. Only users who joined at least one group were used for this experiment.

Table 7: The degree of overlap between the groups joined by pairs of friends does not affect the probability that they use cameras of the same brand, although there does appear to be a small drop for friends having some but only few groups in common

$G_J(X,Y)$	*Mar-May06*	*Mar-May07*	*Mar-May08*
$0 \leq G_J < 0.2$	0.22 (5.3M)	0.25 (9.3M)	0.27 (10M)
$0.2 \leq G_J < 0.4$	0.20 (36K)	0.22 (86K)	0.24 (129K)
$0.4 \leq G_J < 0.6$	0.24 (4.6K)	0.26 (8.8K)	0.26 (9.7)
$0.6 \leq G_J < 0.8$	0.26 (559)	0.24 (1.0K)	0.27 (1.2K)
$0.8 \leq G_J \leq 1.0$	0.26 (3.4K)	0.24 (6.0K)	0.27 (6.8K)

Somewhat surprisingly, Table 7 does not show any correlation between the degree of group similarity and the probability of brand congruence. Interestingly, there appears to be a small drop for friends who "only share very few, but more than no groups". Maybe sharing at least one group is an indication of at least being aware of the other person's groups, whereas not sharing more than one might then be a conscious decision *against* doing so. Not having any groups in common might be simply due to not even looking at the other person's groups joined. Also recall that we only have the group information for the time of crawling, which is only an approximation of the groups joined in 2006 and 2007.

Mutual vs. one-way links. As opposed to most other social networks, such as Facebook or Myspace, Flickr allows one-way friendship links. Such links can serve the purpose of a "bookmark" for the other person's profile or they could be a sign of admiration. The "admired" person is informed about the created link and can then decide to reciprocate it or not. One could expect that reciprocated links are stronger indications for actual friendship and tend to lead to higher brand congruency than one-way links. Or one could hypothesize that one-way links, if they are indications of admiration, would lead to higher brand congruency. To test if there is any difference, we split the friendship links into two classes, mutual links and one-way links. Table 8 shows that, somewhat surprisingly, there are no significant differences between the two types of friendship, possibly as these two effects cancel each other out.

Table 8: Probability of brand congruence for two friends, when the friendship type is broken down into mutual and one-way friendship

	Mar-May06	*Mar-May07*	*Mar-May08*
Mutual	0.22 (3.9M)	0.24 (7.1M)	0.27 (8.0M)
One-way	0.22 (2.0M)	0.25 (3.4M)	0.28 (3.4M)

Many friends vs. few friends. Another dimension to explore is the size of a person's list of friends. One could consider a friendship link from a person with few friends to be more "meaningful" than from a user with dozens of friends. To test if there are such difference between these two categories, we split users according to whether they have more than five friends (this is a "large" user in our terminology) or whether they have less than five friends (this is a "small" user). This was then combined with a breakdown into same or different countries, for user with country information.

Table 9: Brand congruence probabilities for pairs of *friends* when conditioned (i) on the size of the friendship lists of both friends and (ii) on a common or different country of the users. A "small" user is one with up to five friends and a large user has more than five friends. Note that for a pair of small-small friends the congruence probability is greatly increased when they are in the same country (31% for 2008) compared to when they are not (21% for 2008)

Type of pair	*Mar-May06*	*Mar-May07*	*Mar-May08*
small-small			
- all	0.29 (43K)	0.28 (83K)	0.30 (83K)
- same country	0.28 (2.7K)	0.28 (4.1K)	0.31 (3.6K)
- diff. country	0.23 (714)	0.22 (992)	0.21 (745)
small-large			
- all	0.24 (143K)	0.25 (266K)	0.28 (258K)
- same country	0.25 (17K)	0.27 (27K)	0.28 (26K)
- diff. country	0.20 (8.2K)	0.22 (13K)	0.25 (11K)
large-small			
- all	0.22 (272K)	0.24 (480K)	0.26 (453K)
- same country	0.24 (28K)	0.25 (41K)	0.27 (38K)
- diff. country	0.20 (20K)	0.21 (30K)	0.24 (25K)
large-large			
- all	0.22 (5.4K)	0.24 (9.7M)	0.27 (11M)
- same country	0.23 (860K)	0.27 (1.4M)	0.29 (1.4M)
- diff. country	0.21 (1.3M)	0.25 (2.1M)	0.28 (2.1M)

Table 9 nicely shows that the congruence between pairs of friends, who are selective when it comes to adding friends on Flickr, is far stronger than between pairs of friends who have more than five friends. Furthermore, for pairs of "small" users the congruence is dramatically influenced by the fact

Table 10: Brand congruence probabilities for pairs of *random* users when conditioned (i) on the size of the friendship lists of both users (who will most likely *not* be friends) and (ii) on a common or different country of the users. Two small, random users are the *least* likely to be congruent, whereas two small friends are the *most* likely to be congruent

Type of pair	*Mar-May06*	*Mar-May07*	*Mar-May08*
small-small			
- all	0.16 (657K)	0.16 (1.3M)	0.18 (1.1M)
- same country	0.17 (6.0K)	0.18 (9.2K)	0.21 (7.2K)
- diff. country	0.15 (25K)	0.15 (41K)	0.17 (31K)
small-large			
- all	0.16 (498K)	0.17 (1.0M)	0.19 (922K)
- same country	0.18 (10K)	0.21 (16K)	0.23 (14K)
- diff. country	0.16 (43K)	0.17 (76K)	0.19 (66K)
large-small			
- all	0.16 (497K)	0.17 (1.0M)	0.19 (921K)
- same country	0.19 (10K)	0.21 (16K)	0.22 (14K)
- diff. country	0.16 (43K)	0.17 (76K)	0.19 (65K)
large-large			
- all	0.17 (376K)	0.19 (766K)	0.21 (784K)
- same country	0.21 (17K)	0.24 (29K)	0.25 (28K)
- diff. country	0.17 (76K)	0.19 (141K)	0.21 (140K)

whether the two friends are in the same country or not. Table 10 shows the comparison results for random users of the same/different country, broken down the results into the sizes of each users. This shows that "small" users do not inherently have a strong tendency to share the same brand, even if they are in the same country, but that this effect is clearly related to the friendship. Also observe that Table 9 shows that if a pair of small-small friends were both already active on flickr in 2006, then they have a much higher probability of being congruent in 2006, than two large users in the same year (29% vs. 22%). Similarly, the difference between pairs of small users and pairs of large users seems to be getting weaker, which is probably an indication for the fact that the *age* of a friendship link, i.e., when the friendship was established, seems to be an important factor. Unfortunately, this information cannot be obtained directly by us and would have to be traced over a long period.

Varying degree of "cliqueness". Apart from "selectivity" one might also expect that the degree of "cliqueness" between two friends will have an effect. If two friends have many friends in common, they can be said to be in a social clique. Formally, we define the *cliqueness* F_J between between two users X and Y to be the Jacquard coefficient of their two sets of friends. That is, for a pair of friends (X, Y) we have $F_J(X,Y) = |\bar{F}(X) \cap \bar{F}(Y)|/|\bar{F}(X) \cup \bar{F}(Y)|$. Here, $\bar{F}(X) = F(X) \cup \{X\}$, which is the set of X's friends together with X itself. The cliqueness between two users is 1.0 if and only if they share all their friends (which includes the case where they do not have any friends apart from each other[16]), and it can never be zero.

Figure 4 shows how an increase in cliqueness (on the X-axis) leads to a higher brand congruence. Table 11 gives a

[16]But this case is not included in our dataset, as we would have never discovered this isolated component in our crawl.

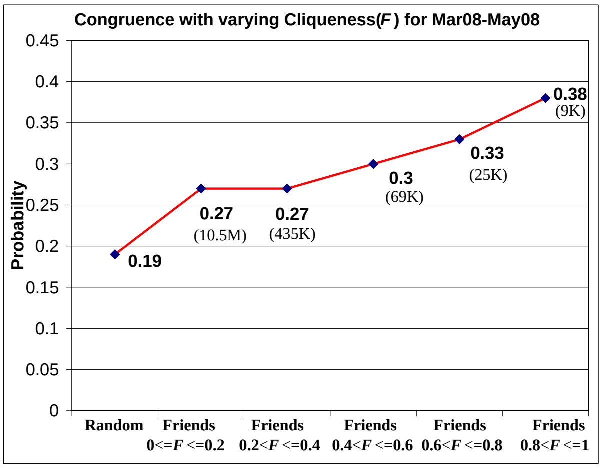

Figure 4: Congruence probability for pairs of friends with varying degree of cliqueness. Cliqueness turned out to be the strongest camera-type independent factor in our study.

Table 11: Breakdown of the congruence probabilities between for friends according to a common/different country and the degree of cliqueness for the two users

$F_J(X,Y)$	*Mar-May06*	*Mar-May07*	*Mar-May08*
$0 \leq F_J \leq 0.5$			
- all	0.22 (5.9M)	0.24 (10M)	0.27 (11M)
- same country	0.23 (906K)	0.27 (1.4M)	0.29 (1.5M)
- diff. country	0.21 (1.3M)	0.24 (2.1M)	0.28 (2.1M)
$0.5 < F_J \leq 1.0$			
- all	0.32 (27K)	0.32 (51K)	0.33 (53K)
- same country	0.30 (2.0K)	0.34 (2.8K)	0.33 (2.5K)
- diff. country	0.28 (317)	0.22 (383)	0.24 (346)

breakdown of users according to low (≤ 0.5) and high (> 0.5) cliqueness in combination with conditioning on a common or different country. Here one can see, by looking at the absolute numbers of links used in each cell, that high cliqueness friends have a far higher probability of being in the same country (91%) than low cliqueness friends (42%), given that both friends provided country information. More interestingly, as already observed for the breakdown into small and large friends, there seems to be a dependence on time, such that the difference between high and low cliqueness friends seems to be disappearing. Again, as we could only observe the present status of a friendship relation, it would be interesting to look at this more closely by taking the *age* of friendship links into account.

When the size of the friendship lists is combined with the degree of cliqueness, an interesting reversal takes places. Whereas, globally, pairs of small users are more likely to be congruent than pairs of large users, Table 12 show this order is reversed, when we only look at friends with a high degree of cliqueness. This can be understood by observing that sharing more than 50% of friends, when both people involved have long lists of friends, is arguably a stronger sign of a clique than if both sides only have a single friend, but this friend is in common. In fact, if we require the cliqueness

Table 12: A more detailed breakdown of the congruence probabilities for 2008 according to (i) the degree of cliqueness F_J, (ii) whether two friends share a common country, and according to (iii) the size of the friendship lists of the two friends. Note that neither the counts for the same and different countries have to add up to "all" (as only users with country information are used for the first), nor do the numbers for small-small and large-large have to add up to "all" (as there are also other edge types)

	User type (more or less than five friends)		
$F_J(X,Y)$	all	small-small	large-large
$0 \leq F_J \leq 0.5$			
- all	0.27 (11M)	0.29 (58K)	0.27 (11M)
- same country	0.29 (1.5M)	0.30 (2.7K)	0.29 (1.4M)
- diff. country	0.28 (2.1M)	0.20 (569)	0.28 (2.1M)
$0.5 < F_J \leq 1.0$			
- all	0.33 (53K)	0.32 (26K)	0.35 (22K)
- same country	0.33 (2.5K)	0.33 (945)	0.32 (1.2K)
- diff. country	0.24 (346)	0.24 (176)	0.27 (32)

to be larger than 0.8 (and not just 0.5), then pairs of large friends have a probability of 48% to share a common brand in 2008 (out of 2,268 pairs). For 2007, this percentage goes up to 51% (out of 2,740 pairs) and it is even 57% (out of 1,365 pairs) in 2006.

Regular vs. sporadic users. We also differentiated between users who have uploaded in total (i) more or (ii) less than 200 public images. This way we hoped to see, if an increase in brand congruence between two random users or two friends, could be related to a similar level of (upload) activity on Flickr. Overall, two more regular users were indeed more likely to be congruent (29% for friends and 23% for random users in 2008) than two more sporadic users (25% for friends and 17% for random users). However, this order is *reversed* when we look only at high-cliqueness pairs of friends. Among these, 35% (out of 30k) were congruent if both friends were sporadic users, compared to 32% (out of 7.4K) if both friends were regular users.

Two important user types: P&S vs. DSLR. Apart from the social aspects studied above, we also looked at a breakdown of "low-end vs. high-end" users. Concretely, we focused on those users who either used a point-and-shoot camera (P&S), which is usually a cheaper (by comparison) camera for non-expert users and on those with a digital single-lens reflex camera (DSLR), which is usually a more expensive camera for more advanced users. The results for friends are presented in Table 13 and in Table 14 for random users.

There are some interesting and surprising observations to make. First, for pairs of P&S users the effect of being in the same country is striking. For two friends in this group, the difference for 2008 is between 29% (for the same country) and 21% (for different countries). For two random users the corresponding numbers are 28% vs. 18%. This even goes so far that the friendship information seems nearly irrelevant in this setting and it shows that the local dominance of "cheap" camera brands differs around the globe. It also seems to imply that users of such cheap P&S cameras are not strongly affected by the influence of their friends. However, this later claim will be refuted, once we additionally take cliqueness into account (Table 15). Second, pairs of DSLR users are

Table 13: Breakdown of congruence probabilities for pairs of friends according to the camera quality used (point-and-shoot vs. digital-single-lens-reflex) and further conditioning on the country

Type of pair	*Mar-May06*	*Mar-May07*	*Mar-May08*
P&S - P&S			
- all	0.23 (1.5M)	0.25 (2.1M)	0.26 (1.7M)
- same country	0.26 (232K)	0.28 (294K)	0.29 (226K)
- diff. country	0.18 (280K)	0.19 (351K)	0.21 (240K)
P&S - DSLR			
- all	0.18 (663K)	0.19 (1.3M)	0.20 (1.3M)
- same country	0.22 (109K)	0.23 (183K)	0.24 (180K)
- diff. country	0.16 (157K)	0.17 (28K)	0.18 (266K)
DSLR - P&S			
- all	0.18 (580K)	0.19 (1.1M)	0.20 (1.1M)
- same country	0.23 (96K)	0.24 (165K)	0.25 (168K)
- diff. country	0.15 (131K)	0.16 (234K)	0.18 (225K)
DSLR - DSLR			
- all	0.49 (447K)	0.48 (1.3M)	0.47 (2.1M)
- same country	0.52 (87K)	0.50 (231K)	0.49 (338K)
- diff. country	0.47 (110K)	0.45 (316K)	0.46 (470K)

far more likely to use cameras of the same brand than other users. Third, for such pairs of "expert" users, the country influence is much weaker and the global differences seem to be washed away for high-end models. These users also seem to show a higher influence to peer pressure, as the degree of congruence, though already at a high level for random users, increases further when conditioned on friendship (42% vs. 47% for 2008, 45% vs. 49% for 2008 and a common country).

Table 14: Breakdown of congruence probabilities for pairs of random users according to the camera quality used (point-and-shoot vs. digital-single-lens-reflex) and further conditioning on the country

Type of pair	*Mar-May06*	*Mar-May07*	*Mar-May08*
P&S - P&S			
- all	0.19 (652K)	0.19 (1.2M)	0.20 (799K)
- same country	0.25 (14K)	0.26 (19K)	0.28 (13K)
- diff. country	0.18 (55K)	0.18 (82K)	0.18 (52K)
P&S - DSLR			
- all	0.17 (167K)	0.17 (434K)	0.19 (454K)
- same country	0.25 (5K)	0.26 (10K)	0.27 (10K)
- diff. country	0.15 (16K)	0.18 (82K)	0.17 (38K)
DSLR - P&S			
- all	0.17 (167K)	0.17 (437K)	0.19 (456K)
- same country	0.24 (5K)	0.26 (11K)	0.27 (10K)
- diff. country	0.15 (17K)	0.16 (37K)	0.17 (38K)
DSLR - DSLR			
- all	0.44 (43K)	0.42 (158K)	0.42 (259K)
- same country	0.45 (1.8K)	0.45 (6K)	0.45 (8K)
- diff. country	0.43 (4.9K)	0.41 (16K)	0.41 (27K)

The effect of friendship becomes much more pronounced, when we look at the results for high-cliqueness friends in Table 15. Now a pair of cliqued P&S friends has a probability of 39% of being congruent, compared to 28% for two random P&S users in the same country. Similarly, the congruence goes up to 66% for pairs of highly cliqued DSLR users, compared to 45% for two such random users in the same country and compared to 49% for two such friends (regardless of cliqueness) in the same country. The same table also shows, in agreement with the results in Tables 13 and 14, that for pairs of different user types, any friendship connection is essentially irrelevant! This was at least in so far surprising to us, as we were expecting the "inexperienced" P&S users to be influenced by their "expert" DSLR user friends, which does not seem to be the case.

Table 15: Breakdown of congruence probabilities for pairs of friends for the period March-May 2008 according to the camera quality used and the degree of cliqueness ($F_J \leq 0.5$ or $F_J > 0.5$)

	Cliqueness		
Type of pair	ignored	low	high
all	0.27 (11M)	0.27 (11M)	0.33 (53K)
P&S - P&S	0.26 (1.7M)	0.26 (1.7M)	0.39 (15K)
P&S - DSLR	0.20 (1.3M)	0.20 (1.3M)	0.23 (4.1K)
DSLR - P&S	0.20 (1.1M)	0.20 (1.1M)	0.23 (4.1K)
DSLR - DSLR	0.47 (2.1M)	0.47 (2.1M)	0.66 (5.8K)

Evolution over time. So far, our analysis has mostly focused on a single time slot at a time, though we did hint at some effects apparently due to the age of friendship links. In this and the next paragraphs, we will focus on the evolution of brand congruence more explicitly. The histogram in Figure 5 shows how pairs of users have become more and more likely to share a brand. For pairs of friends this effect is at least partly due to the fact that the current friendship graph is only an approximation of the friendships in 2007 or 2006. But as this effect exists even for pairs of random users, it shows that the global variance in terms of brand distribution has gone down. More concretely, the two strongest brands (Canon and Sony) combined, increased their share of users from 48% in 2006, to 49% in 2007 and finally to 54% in 2008 and the entropy of the global brand distribution has decreased from 3.1 bits in 2006 to 2.9 bits in 2008.

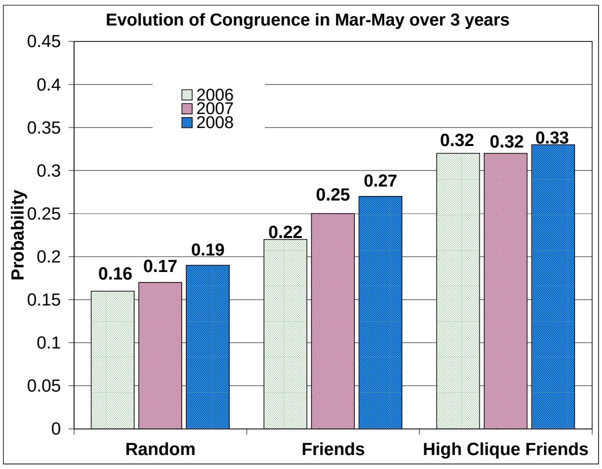

Figure 5: Changes in congruence probabilities over time. Recall that all friendship links are with respect to the network in 2008.

Loyalty for low-budget vs. high-end users Intuitively, one would expect photographers to "evolve" in terms of brand loyalty. That is, in their early days as photographers they would be using comparably cheap point-and-shoot cameras and they would not feel particularly attached to their current brand. For these users the buying decision was probably more influenced by offers at major electronic discount stores, than it was by consciously considering all possible options in terms of their quality. Eventually, they might decide to take photography more seriously and purchase a single lens reflex camera. At this point one would expect their decision to be more conscious and the user feeling more attached to the particular model he purchased. This would then lead to a higher brand loyalty when it comes to future purchases.

This is indeed what we observed. Namely, users of point-and-shoot cameras are far more likely to switch to a different brand when they buy a new camera, then users of single-lens-reflex cameras. See Table 16 for details. However, this has to be put in relation to the observation that DSLR users *generally* have less variability in terms of their brands used and tend to stick to fewer brands. Still, comparing the numbers of Table 16 with Table 14 one observes that a random DSLR user in the same country "only" has a probability of 45% of sharing my brand in 2008, but if I change my model from 2007 to 2008, and if I have a DSLR camera in 2007, then I have a probability of 64% of ending up with the same brand again. The corresponding numbers are 29% and 48% for P&S users. Somewhat surprisingly though, the probability of staying with the same brand is in both cases *lower* than the probability of sharing the same brand with a high cliqueness friend of the same user type (DSLR or P&S). This is probably due to the fact that the event of a user changing her model is already a small indication of dissatisfaction, which might encourage a brand change.

Table 16: Users of DSLR cameras are, compared to users of point-and-shoot cameras, (i) less likely to change their camera model over the course of a year and (ii) less likely to change their brand, even if they do change their camera model. The last statement remains true when the fact that there is generally less variability among DSLR users is taken into account (Tables 13 and 14). Unrelated to the brand loyalty aspect, it is also noteworthy that more than half (!) of the P&S users changed their model between 2007 and 2008

	Camera type in 2007	
	P & S	DSLR
Number of users considered	142K	66K
Changes in brand	34%	15%
Changes in model	52%	36%
Brand change, given model change	67%	40%

Triggering of events by friends. We also tried to investigate, whether the fact that a user changes her brand or at least acquires a new camera model, has some measurable influence on the probability that her friends do likewise, especially the high cliqueness friends, with a cliqueness higher than 0.5. Again, such effects are indeed observable.

Given a user changes her camera model between 2007 and 2008, on average 54% of her high cliqueness friends and 51% of her low cliqueness friends also change their model. However, a random user only has a probability of 48% of doing so. Also, out of all the low cliqueness friends who change their model, 29% change to a camera of the same brand as the user of attention, and this percentage increases to 38% for the cliqueness friends. This percentage is higher than the 33% probability of two high cliqueness friends in 2008 to share a common brand, which is a further indication that they indeed changed together. On the other hand, random users who change their model during the same period, only have a probability of 20% of changing to the same brand in 2008. This percentage is then just one percent above the 19% probability for two random user to share a brand.

Similarly, given that a user changes her brand (and hence the model) in the same period, on average 38% of her low cliqueness friends do the same (out of which 18% change to the same brand), 43% of the high cliqueness friends (out of which 27% change to the same brand) and 38% of random users (out of which 13% change to the same brand). In all of the cases above, taking a common country into account changed very little, generally adding 1-2% to all numbers.

5. SUMMARY AND FUTURE WORK

In this work, our analysis was focused on testing (i) whether friendship on Flickr has a significant impact on brand congruence ("Yes."), (ii) whether this impact can be explained by geographical factors alone ("No."), (iii) which factors have the biggest influence on the "strength" of a friendship link regarding brand congruence ("Cliqueness and size of friendship lists."), (iv) which user types need to be considered separately ("DSLR and P&S."), and (v) whether there are noticeable effects of a local *change* of models or brands being related to changes in the neighborhood ("Yes.").

Building on these insights, we plan to dig deeper and, e.g., consider the effect of a common city, look at how brand loyalty differs between different brands and identify individual *key users* with a measurable effect over others, similar to the issues addressed in [3] and [9]. Here, promising factors to investigate are (i) the number of comments left for a user's pictures, (ii) the number of times these pictures are bookmarked by others and (iii) the relative time of a certain model change compared to the time of similar changes in the neighborhood. Similarly, one could try to look at if users leaving comments on other users pictures are more susceptible to influences due the friendship network. This would then go hand in hand with a closer look at the *evolution* of the brands for nodes in the network, in the spirit of [3]. These issues would further strengthen the relevance of our work for targeted advertisements in social networks. It also looks promising to verify our observations for other social networks and other types of products, ranging from books to fashion articles.

6. REFERENCES

[1] Y.-Y. Ahn, S. Han, H. Kwak, S. Moon, and H. Jeong. Analysis of topological characteristics of huge online social networking services. In *16th International Conference on the World Wide Web (WWW)*, pages 835–844, 2007.

[2] R. Algersheimer, U. M. Dholakia, and A. Herrmann. The social influence of brand community: Evidence from european car clubs. *Journal of Marketing*, 69:19 – 34, 2005.

[3] L. Backstrom, D. Huttenlocher, J. Kleinberg, and X. Lan. Group formation in large social networks: membership, growth, and evolution. In *12th International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 44–54, 2006.

[4] A. Chaudhuri and M. B. Holbrook. The chain of effects from brand trust and brand affect to brand performance: The role of brand loyalty. *Journal of Marketing*, 65:81 – 93, 2001.

[5] N. Garg and I. Weber. Personalized, interactive tag recommendation for flickr. In *Conference on Recommender Systems (RecSys)*, pages 67–74, 2008.

[6] J. Kleinberg. Cascading behavior in networks: algorithmic and economic issues. In N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani, editors, *Algorithmic Game Theory*, chapter 24. Cambridge University Press, 2007.

[7] J. Leskovec, L. A. Adamic, and B. A. Huberman. The dynamics of viral marketing. *ACM Transanctions on the Web*, 1(1):5, 2007.

[8] J. Leskovec and E. Horvitz. Planetary-scale views on a large instant-messaging network. In *17th International Conference on the World Wide Web (WWW)*, pages 915–924, 2008.

[9] J. Leskovec, A. Singh, and J. M. Kleinberg. Patterns of influence in a recommendation network. In *Advances in Knowledge Discovery and Data Mining, 10th Pacific-Asia Conference (PAKDD)*, pages 380–389, 2006.

[10] C. Marlow, M. Naaman, D. Boyd, and M. Davis. Ht06, tagging paper, taxonomy, flickr, academic article, to read. In *7th Conference on Hypertext and Hypermedia (HT)*, pages 31–40, 2006.

[11] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee. Measurement and analysis of online social networks. In *7th SIGCOMM Conference on Internet measurement (IMC)*, pages 29–42, 2007.

[12] T. Rattenbury, N. Good, and M. Naaman. Towards automatic extraction of event and place semantics from flickr tags. In *30th International SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*, pages 103–110, 2007.

[13] P. Reingen, B. Foster, J. J. Brown, and S. Seidman. Brand congruence in interpersonal relations: A social network analysis. *Journal of Consumer Research*, 11(3):771–783, 1984.

[14] M. Richardson and P. Domingos. Mining knowledge-sharing sites for viral marketing. In *8th International SIGKDD Conference on Knowledge Discovery and Data Mining (SIGKDD)*, pages 61–70, 2002.

[15] B. Sigurbjornsson and R. van Zwol. Flickr tag recommendation based on collective knowledge. In *17th International Conference on the World Wide Web (WWW)*, pages 327–336, 2008.

[16] P. Singla and M. Richardson. Yes, there is a correlation - from social networks to personal behavior on the web. In *17th International Conference on the World Wide Web (WWW)*, pages 655–664, 2008.

[17] W.-S. Yang, J.-B. Dia, H.-C. Cheng, and H.-T. Lin. Mining social networks for targeted advertising. *Hawaii International Conference on System Sciences (HICSS)*, 6:137a, 2006.